FEBS Letters 585 (2011) 1589–1594

FEBS Letters
journal homepage: www.FEBSLetters.org

Review

Retrotransposons: Mobile and mutagenic from conception to death

Geoffrey J. Faulkner *

Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush EH25 9RG, UK

ARTICLE INFO

Article history:
Received 28 February 2011
Revised 25 March 2011
Accepted 28 March 2011
Available online 6 April 2011

Edited by Sergio Papa, Gianfranco Gilardi and Wilhelm Just

Keywords:
Retrotransposon
L1
Alu
SVA
Genomics
High-throughput sequencing

ABSTRACT

Mobile genetic elements feature prominently in mammalian genome evolution. Several transposition-competent retrotransposon families (L1, *Alu*, SVA) remain active in the human germ line, leading to pathogenesis as well as genome structural variation across the global population. High-throughput screening approaches have recently been developed to detect retrotransposon insertion polymorphisms. Evidence produced by these and other genome-scale technologies indicates an expanded role for retrotransposition in human biology, including somatic mobilisation in the developing embryo and in neural cells.

1. Introduction

Retrotransposons are mobile genetic elements that invade eukaryotic genomes and then spread via a "copy-and-paste" mechanism [1,2]. Retrotransposition is a powerful mutagenic process that can completely alter the complexity, size and structure of a genome over a comparatively brief evolutionary timescale [3]. Specialised biological pathways have evolved to control the phenomenon, particularly in plants and animals, where transposed elements occupy a high percentage of genomic DNA [4–7]. Nearly half of the human genome is derived from retrotransposons, transposons and endogenous retroviruses, with the latter two classes likely now including immobile elements only. Despite an overtly parasitic pattern of replication, these sequences distinguish our evolutionary history from that of other animals. For instance, inactivation of CMP-*N*-acetylneuraminic acid hydroxylase by an *Alu* retrotransposon subsequent to the Homo-Pan divergence is considered to have influenced human brain expansion [8]. Genome structural variation generated by retrotransposons is also common in humans and is linked to ~75 diseases [9–13]. Thus, retrotransposition yields human-specific genetic diversity underlying phenotypic variation.

Three retrotransposon families are currently active in humans: L1, *Alu* and SVA (Fig. 1). L1 is considered the master controller of retrotransposition [14]; its two open reading frames encode proteins [15–17] capable of mobilising L1 RNAs in *cis* as well as a host of other RNAs in *trans*, including those produced by *Alu*, SVA and in some cases protein-coding genes [18–21]. Ostensibly due to the low processivity of the L1 reverse transcriptase, the vast majority of retrotransposon-derived sequences in the genome have been immobilised by truncations, as well as inversions [22] and other mutations. As first hypothesised by McClintock and later developed by Britten and Davidson [2,23], these transposed elements provide raw material for gene regulation circuits. More recent experiments suggest that transposed elements are widely transcribed and only partially repressed by the same epigenetic marks that control their transposition-competent relatives [24–26].

Post-translational modification of histone tails (e.g., acetylation, methylation) and DNA methylation are the primary mechanisms that inhibit retrotransposon mobilisation. In human, *MECP2* associates with its co-repressors *HDAC1*, *HDAC2* and *SOX2* to silence the canonical L1 5′ promoter [27–29]. Removal of this epigenetic repression enables transcription factors, such as *RUNX3* and *YY1* [30,31], to bind to the L1 promoter and direct PolII to produce a full-length L1 RNA. If this RNA can then avoid a second level of less stringent genome surveillance based on the RNAi [32] and RNA editing [33,34] pathways it can be translated to generate the L1 transposition machinery. Acting in *cis*, this complex produces new L1 copies via target-primed reverse transcription (TPRT) [35].

The specific mechanism of retrotransposition varies across families. *Alu* for instance is transcribed from an internal PolIII promoter

* Fax: +44 (0) 131 440 0434.
E-mail address: geoff.faulkner@roslin.ed.ac.uk

0014-5793/$36.00
doi:10.1016/j.febslet.2011.03.061

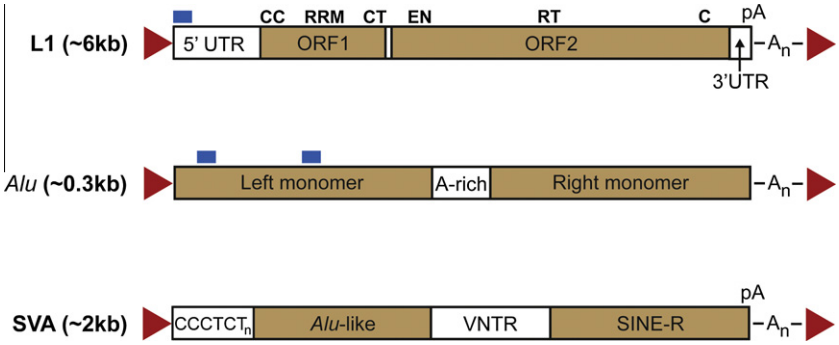

Fig. 1. Active human retrotransposon consensus sequences. L1 is comprised of two open reading frames (ORFs) flanked by 5′ and 3′ untranslated regions (UTRs). ORF1 contains coiled-coil (CC), RNA recognition motif (RRM) and carboxyl-terminal (CT) domains [15]. ORF2 encodes endonuclease (EN), reverse transcriptase (RT) and cysteine-rich (C) domains [16,17]. *Alu* comprises two monomeric regions separated by a short, A-rich A_5TACA_6 consensus sequence. SVA is a composite element combining a hexamer repeat ($CCCTCT_n$) with two SINEs separated by a variable-number-of-tandem repeats (VNTR) region. The 5′ SINE resembles an *Alu* whereas the 3′ SINE is derived from a long terminal repeat (LTR) associated with a human endogenous retrovirus (HERV-K10) [20]. Each element typically generates a target site duplication (red arrows) upon integration and incorporates a poly(A) tail. The 3′ termini of L1 and SVA are marked by a polyadenylation signal (pA). L1 and *Alu* are transcribed from internal PolII and PolIII promoters, respectively (blue boxes). Figure adapted from [1,9].

[36] and SVA likely relies upon external 5′ promoters to initiate transcription by PolII [37]. In some cases multiple "hot" retrotransposons can mobilise from the same family [38] or, as in the case of the rodent *BC1* family, only the RNA of a single master gene can initiate retrotransposition [39]. Despite these variations, virtually all retrotransposition follows the general pattern of (a) escape from epigenetic and post-transcriptional repression, (b) transcription of an RNA intermediate, (c) reverse transcription and (d) integration in a new locus. This review will explore genome-wide evidence for germ line and somatic cell retrotransposition events in the human population.

2. Retrotransposons, genome evolution and disease

Retrotransposition has mainly been observed in germ cells, during very early embryonic development and in cultured carcinoma cells [40–45]. Through studying these systems in depth, we now have a reasonable understanding of the cellular and molecular environments that allow an individual element to mobilise in the human germ line. Earlier experiments also revealed which L1, *Alu* and SVA subfamilies were active in humans compared to other primates [46,47]. Thus, we possess an almost certainly complete catalogue of active human retrotransposon families and a reasonably accurate model of when and how these elements mobilise during development. A major objective is now to map and functionally annotate as many novel integration sites as possible and calculate the frequency of each in the global population.

2.1. High-throughput approaches to map retrotransposon insertions

The full-length nucleotide sequences of L1, *Alu* and SVA were first produced by cloning and sequencing repetitive DNA resistant to endonuclease treatment (*Alu*) [48], through cDNA sequencing (L1) [49] and through sequencing elements found in cloned protein-coding genes (SVA) [20]. The landmark discovery that retrotransposition was still occurring in humans, and indeed could cause disease [50], greatly increased efforts to find recent insertions and characterise retrotransposon subfamilies. However, the catalogue of loci containing active retrotransposons grew slowly until higher-throughput approaches were developed to map retrotransposon integration sites a priori. An important precursor to these approaches, dubbed L1 display, was developed to screen large human cohorts for dimorphic elements [51,52]. As broadly summarised in Fig. 2A, L1 display exploited an ACA trinucleotide diagnostic for the L1-Ta subfamily [38] to detect the 3′ terminus and flanking region of recent L1 insertions. A primer incorporating this ACA was paired with an arbitrary 10mer to generate a PCR amplicon that could be visualised by Southern blot [52]. Dimorphism was confirmed by the absence or presence of a band in a given human population.

L1 display was subsequently modified to incorporate suppression PCR and capillary sequencing in an approached called ATLAS [53]. Loci containing dimorphic L1s could be identified by ATLAS after clones were aligned to an early draft of the human genome sequence produced by whole genome shotgun sequencing (WGSS). In addition to a complete reference genome, which improved the resolution of ATLAS, WGSS also identified numerous retrotransposon insertions of low allelic frequency [4,7,54–57], suggesting that high-throughput methods would be necessary to map rare insertions.

Several groups have recently presented techniques to screen hundreds of individuals for germ line retrotransposition events. Each employed innovative variations to improve the throughput and enrichment for target sequences achieved by ATLAS. Ewing and Kazazian [58] used hemi-specific PCR to amplify the 3′ flanking region of recent L1 insertions, followed by Illumina deep sequencing (Fig. 2B). Beck et al. [59] subjected fosmid libraries to a series of screens, including ATLAS, to isolate the flanking regions of full-length and near full-length L1 sequences (Fig. 2C). Huang et al. [60] applied transposon insertion profiling by microarray (TIP-chip) technology developed in yeast [61] to *Alu* and L1. Focusing on L1, they made substantial modifications to ATLAS by digesting DNA 1-5kb from the L1 3′ terminus (Fig. 2A), followed by ligation-mediated PCR (LM-PCR) and hybridisation to genome-wide tiling arrays (Fig. 2D). Iskow et al. [62] also used restriction endonucleases and LM-PCR to amplify L1 and *Alu* (L1-seq and *Alu*-seq, respectively) but performed capillary and 454 sequencing to locate novel insertions with respect to the reference genome (Fig. 2E).

These technologies massively expanded the available catalogue of L1 and *Alu* polymorphic insertion sites [56]. However, each approach varied considerably in terms of sensitivity, a major consideration if the primary goal was to generate a complete list of retrotransposon integration sites in an individual genome without resorting to WGSS. The method of Beck et al. [59] presented a low false-positive rate and, through screening fosmids for the L1 5′UTR before sequencing, was by design an excellent way to detect hot L1s. A consequential bias towards longer insertions, which are far less common than the short 3′ fragments produced by L1 retrotransposition, and a reliance on capillary sequencing, meant that the approach intentionally detected a minority of L1 insertions.

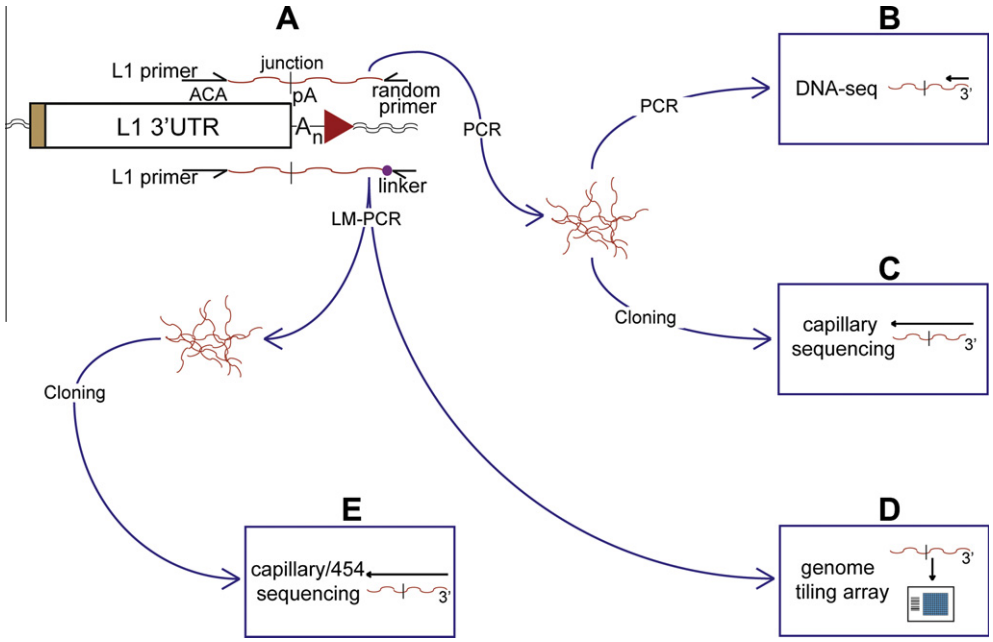

Fig. 2. High-throughput strategies to identify recent L1 insertions. (A) Enrichment for L1 3′ junction sequences [52,53]. An L1 primer incorporating the ACA trinucleotide diagnostic for hot L1s is coupled with a random primer, followed by PCR. Alternatively, amplification is achieved through digesting DNA with one or several restriction enzymes (purple circle) followed by ligation-mediated PCR (LM-PCR). (B) To enable deep sequencing, first round PCR products are further amplified by hemi-specific nested PCR before single-end sequencing with the Illumina platform [58]. (C) Alternatively, first round PCR products are cloned and sequenced by capillary sequencing [59]. (D) Without further amplification, LM-PCR products are hybridised to whole-genome tiling arrays [60]. (E) First round LM-PCR products are cloned and sequenced by capillary or 454 sequencing [62]. Note that the schematic is only representative of each technique. For instance, the method of Beck et al. (C) screens fosmid libraries before enrichment is achieved in (A) [59].

Of the four approaches, TIP-chip was a major methodological outlier in its use of tiling arrays to resolve insertion loci [60]. This simplified sample preparation and data interpretation greatly, meaning that the approach has substantial promise for genotyping. As Huang et al. acknowledge, however, tiling array probes are by necessity designed away from repetitive elements to avoid cross-hybridisation. Given that half of the human genome is comprised of transposed elements, TIP-chip may have failed to detect an equivalent proportion of insertion sites.

Iskow et al. exploited high-throughput sequencing to assay normal and cancer samples for germ line insertions and also made the pivotal observation that somatic L1 retrotransposition occurred in lung tumours [62]. One of the more noticeable and unexplained statistics arising from this data was that, of several hundred thousand 454 reads produced by L1-seq and *Alu*-seq, fewer than 25% could be aligned to the reference genome, despite sequencing from the 3′ end to avoid L1 and *Alu* poly(A) tails. Other than the most obvious explanation of low sequencing quality, this could also be due to the fact that many of the samples used were tumours, where indels and other structural variants could obstruct longer alignments to the reference genome.

Ewing and Kazazian detected a similarly large number of insertions, using fewer samples, through short read (76mer) single-end sequencing [58]. This approach therefore achieved the highest coverage of polymorphic L1 retrotransposition events per individual. Previous results and calculations for short read data [25,63,64] suggest that a 76mer has sufficient specificity to uniquely map to the majority of transposed elements, excluding a small minority of very recent insertions. Paired-end reads could have been used to trace a subset of new insertions back to hot L1s on the genome, but would also have nearly doubled the cost of sequencing. Overall, though it should be emphasised that each of the four methods was designed with substantially different applications in mind, the depth achieved by the Ewing and Kazazian method probably provided the most comprehensive detection of recent L1 insertions in a given sample.

2.2. Retrotransposon insertion polymorphisms in human populations

When combined with large-scale WGSS projects [65], high-throughput screening of retrotransposon insertion sites has provided accurate estimates of L1 and *Alu* activity in the global human population, in addition to preliminary calculations based on a smaller pool of SVA polymorphisms [57]. It is clear that structural and genetic variation caused by retrotransposition is widespread. For *Alu*, comparisons between human individuals and other primates, as well as calculations based on the frequency of disease causing insertions, have produced a consensus of 1 de novo insertion for every 20 live births [55,57,66], a 10-fold increase over estimates from the pre-genome era [10].

By contrast, estimates of L1 retrotransposition rate were revised substantially downward in the past decade. Early calculations based on the frequency of disease causing L1 mutations suggested a rate of 1 insertion per 20 live births [67]. More recently, comparisons between individual genome sequences placed the retrotransposition rate at 1/200 births [57] and the high-throughput screening methods highlighted above again revised this to 1/100–1/150 births [58,60]. As others have suggested, the substantial differences between disease-based and sequencing-based calculations may be due to strong selection against new L1 insertions [9,58].

In terms of transposition competent elements, the human genome on average contains 80–100 L1s, 2000–3000 *Alus* and probably fewer than 100 SVAs, with each presenting a sub pool of hot elements [36,38,68,69]. If these contribute 1 new insertion per 150, 20 and 1000 live births for L1, *Alu* and SVA [57], respectively, the global human population of ~6.8 billion individuals would be expected to contain in the order of 392 million private

retrotransposon insertions (4.5×10^7 L1, 3.4×10^8 *Alu* and 7×10^6 SVA). Prior estimates based on WGSS suggested that any two individuals would likely differ by approximately 180 L1, 1283 *Alu* and 56 SVA polymorphisms [68] and, as most of these occur in more than 1% of the population, deeper screening has not substantially altered these figures [58,60]. Given such vast diversity it is easy to appreciate why high-throughput strategies designed to detect retrotransposon insertions are necessary.

In the future these approaches could be applied to discover the genetic aetiology of diseases in the same manner as a genome-wide association study (GWAS). In a pilot experiment, TIP-chip found intronic L1 insertions in two protein-coding genes of possible functional significance for patients afflicted by one of several X-linked disorders [60]. Though these data were inconclusive, numerous diseases have previously been associated with heritable and de novo retrotransposition events [9,50] and it is likely that high-throughput screening applied to larger patient cohorts would reveal pathogenic polymorphisms. This could be achieved through modified or hybrid versions of each technique discussed above (e.g., L1-seq combined with Illumina sequencing or TIP-chip probes attaining lower cross-hybridisation rates) and will be assisted by ongoing reductions in sequencing and microarray costs.

3. Somatic retrotransposition during neurogenesis

Germ line retrotransposition events are, as noted above, an important source of heritable genetic variation [70]. However, recent assays have suggested that most de novo insertions occur during embryogenesis and, perhaps to a lesser extent, in later development [43,71]. These mutations generate intra-individual variation where genetically distinct somatic cells form a genome mosaic in a particular organ or system [27,72]. Somatic genome mosaicism violates a central tenet of biology: apart from very limited exceptions, such as V(D)J recombination in immunoglobulin genes [73], the normal somatic cells of the human body are considered genetically identical. Widespread genome structural variation caused by retrotransposition during ontogenesis could fundamentally alter the genotype-phenotype paradigm, in a similar fashion to Waddington's epigenetic landscape of the 1950s [74].

Perhaps the best evidence for somatic genome mosaicism arises from L1 retrotransposition during neurogenesis [27,28,75,76]. In this system, epigenetic repression of the L1 promoter is not ubiquitous in neural precursors and committed neural cell lineages [27,28]. Thus, hot L1s can initiate transcription of full-length RNAs [25,77,78] and mobilise to new genomic loci, without the need for cell division [79]. Transposition has been observed directly in cultured neural cells using transgenic L1s tagged with an EGFP reporter cassette that fluoresces upon integration [27,75] and indirectly in vivo by detection of L1 copy number variation (CNV) across adult human brain tissues [28,75].

Despite these compelling experiments, further evidence is required in support of the somatic mosaicism model. Firstly, transgenic L1 retrotransposition rates probably do not perfectly recapitulate levels in vivo. Secondly, the relative level of L1 mobilisation during embryogenesis versus later development – a major factor in producing genetic heterogeneity in adult cells – is not well established. Thirdly, and most importantly, it is unknown which genomic loci harbour somatic L1 insertions in the brain or elsewhere. These data alone could confirm the occurrence of somatic mosaicism and, critically, could be used to link genetic changes, including those in protein-coding regions, with phenotypic effects.

In an attempt to locate somatic L1 retrotransposon events in early embryonic stages, Huang et al. applied TIP-chip to peripheral blood leukocytes from phenotypically discordant monozygotic twins [60]. No somatic events were detected. One explanation for this is that waves of somatic retrotransposition throughout organogenesis could produce vastly different frequencies for each insertion (Fig. 3). Therefore, the majority of retrotransposition events could occur below the detection thresholds of TIP-chip, particularly if further somatic activity in peripheral blood is less common than in the brain. Though Iskow et al. detected somatic L1 retrotransposition in tumours with L1-seq, these mutations could be very common after clonal replication of cancer cells [62]. Mapping the individual retrotransposition events that collectively form a somatic mosaic in a healthy tissue may pose an even more significant challenge to the field.

It is intriguing that retrotransposition could create genetically distinct cell subpopulations in the mammalian brain. The brain is an incredibly complex organ, structurally and functionally, and has well known regenerative and adaptive characteristics related to plasticity and learning [80]. The genetic basis for most neurological disorders is also poorly established [81]. Somatic mosaicism could therefore be of major biological significance [72]. Though this viewpoint is highly speculative, it is plausible that neurodegenerative conditions such as Alzheimer's disease and Parkinson's disease could involve somatic retrotransposition events that dysregulate gene pathways underlying childhood and adult neurogenesis [80]. For instance, recent experiments indicated elevated L1 retrotransposition rates in the brains of Rett syndrome patients [28]. From yet another perspective, a mutagenic system where individual neurons are randomly afflicted by somatic retrotransposon insertions could inactivate tumour suppressor genes as a prelude to tumorigenesis. The relationship between somatic retrotransposition and pathogenesis is currently almost entirely unexplored; it equally possible that somatic retrotransposition is necessary for normal biological processes or indeed has negligible phenotypic effects.

Another open question is whether somatic L1 retrotransposition would follow the same pattern as seen in the germ line, where mutations in exons and introns are negatively selected [4]. DNA

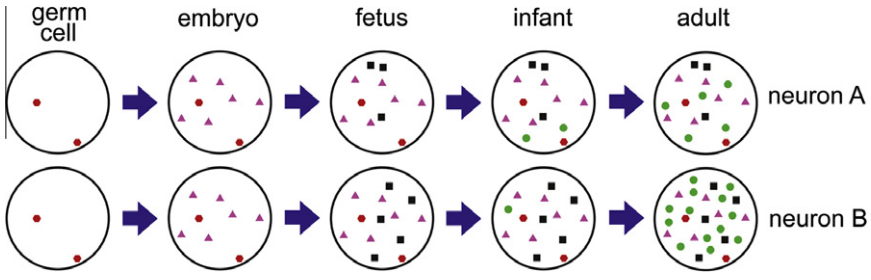

Fig. 3. Schematic representation of germ line and somatic retrotransposition events during the development of two neurons from different brain regions. Germ cell (red hexagon), embryonic (purple triangle), fetal (black squares) and post-natal (green circles) mutations are present in a somatic genome mosaic. Symbols in different locations represent different mutations. Developmental progression is correlated with an increased probability of a retrotransposon insertion being dimorphic.

in open chromatin is known to be more susceptible to nicking by the L1 endonuclease during TPRT [82]. If, for instance, a gene is highly transcribed in neurons, implying an open chromatin state, it would probably also be predisposed to retrotransposon insertions that are not individually selected against during evolution. Also, as others have shown, retrotransposons do not have to insert in exons to have major functional consequences. Mutations in introns, promoters, distal enhancers and other non-coding regions can have pronounced effects [25,83–85]. In this way, somatic retrotransposition could subtly alter the genetic and transcriptional landscape of a cell to better suit a particular environmental niche and, though these mutations would not be inherited per se, the mosaic execution programme could be transmitted through the germ line.

4. Conclusion

Genome-wide technologies have been developed to assess retrotransposition events in germ and somatic cells, indicating major inter- and intra-individual genetic diversity in humans. It is likely that the same evolutionary arms race between retrotransposons and their host genomes that has occurred over hundreds of millions of years will continue to produce structural variation in the human genome. However, a divide remains between the detection of retrotransposon insertions and linking these with phenotypic consequences, particularly for somatic mobilisations. The challenge in this era of personalised genomics [65] will be to place retrotransposon-derived variability in a functional context that informs our basic understanding of healthy development and pathogenesis.

Acknowledgements

The author thanks Kenneth Baillie for editorial suggestions and is supported by a New Investigator award (BB/H005935/1) and an Institute Strategic Programme Grant from the British BBSRC, a C.J. Martin Overseas Based Biomedical Fellowship (575585) from the Australian NHMRC and an EU FP7 Collaborative Project research grant (259743).

References

[1] Kazazian Jr., H.H. (2004) Mobile elements: drivers of genome evolution. Science 303, 1626–1632.
[2] McClintock, B. (1956) Controlling elements and the gene. Cold Spring Harb. Symp. Quant. Biol. 21, 197–216.
[3] Han, K., Xing, J., Wang, H., Hedges, D.J., Garber, R.K., Cordaux, R. and Batzer, M.A. (2005) Under the genomic radar: the stealth model of Alu amplification. Genome Res. 15, 655–664.
[4] Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.
[5] Li, W., Zhang, P., Fellers, J.P., Friebe, B. and Gill, B.S. (2004) Sequence composition, organization, and evolution of the core Triticeae genome. Plant J. 40, 500–511.
[6] Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272–285.
[7] Venter, J.C. et al. (2001) The sequence of the human genome. Science 291, 1304–1351.
[8] Chou, H.H. et al. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc. Natl. Acad. Sci. USA 99, 11736–11741.
[9] Cordaux, R. and Batzer, M.A. (2009) The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. 10, 691–703.
[10] Deininger, P.L. and Batzer, M.A. (1999) Alu repeats and human disease. Mol. Genet. Metab. 67, 183–193.
[11] Kaneko, H. et al. (2011) DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature 471, 325–330.
[12] Ostertag, E.M., Goodier, J.L., Zhang, Y. and Kazazian Jr., H.H. (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. Am. J. Hum. Genet. 73, 1444–1451.
[13] Ostertag, E.M. and Kazazian Jr., H.H. (2001) Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. 35, 501–538.
[14] Kazazian Jr., H.H. and Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. Nat. Genet. 19, 19–24.
[15] Doucet, A.J. et al. (2010) Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet. 6.
[16] Feng, Q., Moran, J.V., Kazazian Jr., H.H. and Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87, 905–916.
[17] Mathias, S.L., Scott, A.F., Kazazian Jr., H.H., Boeke, J.D. and Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. Science 254, 1808–1810.
[18] Dewannieux, M., Esnault, C. and Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35, 41–48.
[19] Esnault, C., Maestre, J. and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24, 363–367.
[20] Shen, L. et al. (1994) Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon–intron structure, composite retroposon, and breakpoint of gene duplication. J. Biol. Chem. 269, 8466–8476.
[21] Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D. and Moran, J.V. (2001) Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439.
[22] Ostertag, E.M. and Kazazian Jr., H.H. (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res. 11, 2059–2065.
[23] Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. Science 165, 349–357.
[24] Faulkner, G.J. and Carninci, P. (2009) Altruistic functions for selfish DNA. Cell Cycle 8, 2895–2900.
[25] Faulkner, G.J. et al. (2009) The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. 41, 563–571.
[26] Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. and Jenuwein, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 24, 800–812.
[27] Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V. and Gage, F.H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435, 903–910.
[28] Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K. and Gage, F.H. (2010) L1 retrotransposition in neurons is modulated by MeCP2. Nature 468, 443–446.
[29] Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389.
[30] Athanikar, J.N., Badge, R.M. and Moran, J.V. (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res. 32, 3846–3855.
[31] Yang, N., Zhang, L., Zhang, Y. and Kazazian Jr., H.H. (2003) An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res. 31, 4929–4940.
[32] Yang, N. and Kazazian Jr., H.H. (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat. Struct. Mol. Biol. 13, 763–771.
[33] Mattick, J.S. (2010) RNA as the substrate for epigenome–environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition. Bioessays 32, 548–552.
[34] Schumann, G.G. (2007) APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem. Soc. Trans. 35, 637–642.
[35] Cost, G.J., Feng, Q., Jacquier, A. and Boeke, J.D. (2002) Human L1 element target-primed reverse transcription in vitro. EMBO J. 21, 5899–5910.
[36] Comeaux, M.S., Roy-Engel, A.M., Hedges, D.J. and Deininger, P.L. (2009) Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? Genome Res. 19, 545–555.
[37] Hancks, D.C., Ewing, A.D., Chen, J.E., Tokunaga, K. and Kazazian Jr., H.H. (2009) Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 19, 1983–1991.
[38] Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian Jr., H.H. (2003) Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. USA 100, 5280–5285.
[39] DeChiara, T.M. and Brosius, J. (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc. Natl. Acad. Sci. USA 84, 2624–2628.
[40] Garcia-Perez, J.L., Doucet, A.J., Bucheton, A., Moran, J.V. and Gilbert, N. (2007) Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res. 17, 602–611.
[41] Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S. and Moran, J.V. (2007) LINE-1 retrotransposition in human embryonic stem cells. Hum. Mol. Genet. 16, 1569–1577.
[42] Garcia-Perez, J.L. et al. (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature 466, 769–773.
[43] Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M. and Kazazian Jr., H.H. (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. 23, 1303–1312.
[44] Macia, A. et al. (2011) Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol. Cell. Biol. 31, 300–316.

[45] Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian Jr., H.H. (1996) High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927.
[46] Salem, A.H. et al. (2003) Alu elements and hominid phylogenetics. Proc. Natl. Acad. Sci. USA 100, 12787–12791.
[47] Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A. and Batzer, M.A. (2005) SVA elements: a hominid-specific retroposon family. J. Mol. Biol. 354, 994–1007.
[48] Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. and Schmid, C.W. (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J. Mol. Biol. 151, 17–33.
[49] Skowronski, J., Fanning, T.G. and Singer, M.F. (1988) Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. 8, 1385–1397.
[50] Kazazian Jr., H.H., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G. and Antonarakis, S.E. (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332, 164–166.
[51] Seleme, M.C., Vetter, M.R., Cordaux, R., Bastone, L., Batzer, M.A. and Kazazian Jr., H.H. (2006) Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. Proc. Natl. Acad. Sci. USA 103, 6611–6616.
[52] Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A. and Swergold, G.D. (2000) Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res. 10, 1496–1508.
[53] Badge, R.M., Alisch, R.S. and Moran, J.V. (2003) ATLAS: a system to selectively identify human-specific L1 insertions. Am. J. Hum. Genet. 72, 823–838.
[54] Ewing, A.D. and Kazazian Jr., H.H. (2011) Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res. 20, 1262–1270.
[55] Hormozdiari, F. et al. (2010) Alu repeat discovery and characterization within human genomes. Genome Res.
[56] Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M.A. and Liang, P. (2006) DbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. Hum. Mutat. 27, 323–329.
[57] Xing, J. et al. (2009) Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 19, 1516–1526.
[58] Ewing, A.D. and Kazazian Jr., H.H. (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 20, 1262–1270.
[59] Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M. and Moran, J.V. (2010) LINE-1 retrotransposition activity in human genomes. Cell 141, 1159–1170.
[60] Huang, C.R. et al. (2010) Mobile interspersed repeats are major structural variants in the human genome. Cell 141, 1171–1182.
[61] Wheelan, S.J., Scheifele, L.Z., Martinez-Murillo, F., Irizarry, R.A. and Boeke, J.D. (2006) Transposon insertion site profiling chip (TIP-chip). Proc. Natl. Acad. Sci. USA 103, 17632–17637.
[62] Iskow, R.C. et al. (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141, 1253–1261.
[63] Faulkner, G.J., Forrest, A.R., Chalk, A.M., Schroder, K., Hayashizaki, Y., Carninci, P., Hume, D.A. and Grimmond, S.M. (2008) A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. Genomics 91, 281–288.
[64] Hashimoto, T., de Hoon, M.J., Grimmond, S.M., Daub, C.O., Hayashizaki, Y. and Faulkner, G.J. (2009) Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRescueLite. Bioinformatics 25, 2613–2614.
[65] Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.
[66] Cordaux, R., Hedges, D.J., Herke, S.W. and Batzer, M.A. (2006) Estimating the retrotransposition rate of human Alu elements. Gene 373, 134–137.
[67] Kazazian Jr., H.H. (1999) An estimated frequency of endogenous insertional mutations in humans. Nat. Genet. 22, 130.
[68] Bennett, E.A., Coleman, L.E., Tsui, C., Pittard, W.S. and Devine, S.E. (2004) Natural genetic variation caused by transposable elements in humans. Genetics 168, 933–951.
[69] Mills, R.E., Bennett, E.A., Iskow, R.C. and Devine, S.E. (2007) Which transposable elements are active in the human genome? Trends Genet. 23, 183–191.
[70] Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D. and Knowles, B.B. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell 7, 597–606.
[71] van den Hurk, J.A. et al. (2007) L1 retrotransposition can occur early in human embryonic development. Hum. Mol. Genet. 16, 1587–1592.
[72] Singer, T., McConnell, M.J., Marchetto, M.C., Coufal, N.G. and Gage, F.H. (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci. 33, 345–354.
[73] Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978) A complete immunoglobulin gene is created by somatic recombination. Cell 15, 1–14.
[74] Waddington, C.H. (1957) The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology, Allen & Unwin, London.
[75] Coufal, N.G. et al. (2009) L1 retrotransposition in human neural progenitor cells. Nature 460, 1127–1131.
[76] Kuwabara, T. et al. (2009) Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nat. Neurosci. 12, 1097–1105.
[77] Belancio, V.P., Roy-Engel, A.M., Pochampally, R.R. and Deininger, P. (2010) Somatic expression of LINE-1 elements in human tissues. Nucleic Acids Res. 38, 3909–3922.
[78] Rangwala, S.H., Zhang, L. and Kazazian Jr., H.H. (2009) Many LINE1 elements contribute to the transcriptome of human somatic cells. Genome Biol. 10, R100.
[79] Kubo, S., Seleme, M.C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian Jr., H.H. and Kasahara, N. (2006) L1 retrotransposition in nondividing and primary human somatic cells. Proc. Natl. Acad. Sci. USA 103, 8036–8041.
[80] Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A.M., Nordborg, C., Peterson, D.A. and Gage, F.H. (1998) Neurogenesis in the adult human hippocampus. Nat. Med. 4, 1313–1317.
[81] Bertoli-Avella, A.M., Oostra, B.A. and Heutink, P. (2004) Chasing genes in Alzheimer's and Parkinson's disease. Hum. Genet. 114, 413–438.
[82] Cost, G.J., Golding, A., Schlissel, M.S. and Boeke, J.D. (2001) Target DNA chromatinization modulates nicking by L1 endonuclease. Nucleic Acids Res. 29, 573–577.
[83] Bejerano, G. et al. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90.
[84] Han, J.S., Szak, S.T. and Boeke, J.D. (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 429, 268–274.
[85] Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A. and Mager, D.L. (2007) Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet. 3, e10.